Optimal Combining Data for Improving Ocean Modeling

L.I.Piterbarg
University of Southern California
Center of Applied Mathematical Sciences
Kaprielian Hall, R108, 3620 Vermont Avenue
Los Angeles, CA 90089-2532
Phone (213) 740 2459, fax (213) 740 2424, e-mail piter@usc.edu
Award # N00014-11-1-0369
http://www.onr.navy.mil/sci_tech/ocean/onrpgahm.htm

LONG-TERM GOALS

The long range scientific goals of the proposed research comprise: (1) developing rigorous approaches to optimal combining different kinds of observations (images, CTD, HFR, glider, drifters etc) with output of regional circulation models for accurate estimating the upper ocean velocity field, subsurface thermohaline structure, and mixing characteristics (2) constructing computationally efficient and robust estimation algorithms based on alternative parameterizations of uncertainty and comprehensive testing them on synthetic data (3) processing real data in the Adriatic and Ligurian Sea via new techniques

OBJECTIVES

The objectives for the second year of research were:
- Developing and testing methods for fusing glider data with model output and/or ship CTD data.
- Testing radar/drifter data fusion in the framework of twin experiments with a high resolution circulation model and on real data
- Combining radar data with tracer observations (SST, color) for estimating surface velocities with focus on data compatibility
- Estimating finite-size Lyapunov exponent by combining real data and model output
- Further developing theoretical approaches based on fuzzy logic to estimating oceanic parameters from small biased samples.

APPROACH

We develop theoretical approaches to the data fusion problem in context of the possibility theory (fuzzy logic) and in the framework of the classical theory of random processes and fields covered by stochastic partial differential equations. We also design computational algorithms derived from the theoretical findings. A significant part of the algorithm validation is their testing via Monte Carlo simulations. Such an approach provides us with an accurate error analysis. Together with my collaborators from Rosenstiel School of Marine and Atmospheric Research (RSMAS), Consiglio Nazionale delle Ricerche (ISMAR, LaSpezia, Italy), University of Toulon (France), Observatoire Oceanologique de Villefranche sur Mer (France), and Naval Postgraduate School (Monterrey, CA) we implement the algorithms in concrete ocean models such as HYCOM, NCOM, MFS, and NEMO as well as carry out statistical analysis of real data sets by means of new methods.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **30 SEP 2012**	2. REPORT TYPE	3. DATES COVERED **00-00-2012 to 00-00-2012**

4. TITLE AND SUBTITLE **Optimal Combining Data for Improving Ocean Modeling**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Southern California,Center of Applied Mathematical Sciences,Kaprielian Hall, R108, 3620 Vermont Avenue,Los Angeles,CA,90089-2532**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**7**	

WORK COMPLETED

1. Developing and testing methods for fusing glider data with model output and/or ship CTD data.

The problem of combining glider data with model output and/or CTD profiles exposes new challenges in the data fusion business such as a different space resolution (for glider measurements it is essentially higher than for typical models or CTD measurements), specific features of a glider path and uncertainty in its position, and, probably most important, superimposing space and time variability in the case of flows varying in time faster than the glider moves.

During the current period we focused on two issues. First, combining glider data with a model output for steady thermohaline patterns and, second, separating space and time variability in glider observations for fast changing thermohaline structures (etc mesoscale fronts) by attracting available ship CTD observations.

As for the first problem we continued to test the algorithm developed on the first stage using synthetic data. That algorithm is based on the fuzzy logic approach [1-4] and the goal was to compare it with more traditional procedures such as weighted mean (linear interpolation) with respect to the glider downcast frequency k and the model bias ε.

To solve the second problem we have developed and tested three different procedures. The first one included a parameterization of thermohaline patterns following up an estimation of parameters from glider and CTD data. The second algorithm involved a fuzzy regression approach for optimal combining glider and CTD data. Finally, a traditional polynomial regression was proposed for processing glider data with CTD observations serving as a control sample.

All three approaches have been tested on synthetic data and real data in Ligurian Sea

2. Testing radar/drifter data fusion in the framework of twin experiments with a high resolution circulation model and on real data.

In now days measurements by HF radars play more and more important role in investigating surface circulation patterns in coastal regions. The problem is complicated by possible failures of the devices and not sufficient accuracy. Different experiments with direct velocity measurements have shown significant errors: around 15 cm/s (along the North Carolina Coast, 1997 [5]), 7-19 cm/s (along the California Coast, 2004, [5]), 6.6-11.3 cm/s (Korea/Tsushima Strait, 2006, [6]), and 6-13 cm/s (East Coast of Korea, 2007, [7]). In the last paper it was noticed that the accuracy of estimating zonal and meridional components are essentially different. Moreover, appropriate velocity estimates in a certain area can be obtained only if the area is covered at least by two radars. If only one of them works properly, then an important problem arises how to use other available information to restore the surface velocities.

In such a situation a presence of drifters in the area of interest could help. A procedure constructed earlier for combining drifter observations with radar data now has been tested first via twin experiments with NEMO model (OPA) and then applied to real data on the Var Coast (Mediterranean).

3. Combining radar data with tracer observations (SST, color) for estimating surface velocities with focus on data compatibility

In the aforementioned situation with only one working radar SST or color observations can be used to retrieve surface velocities. A fuzzy logic based algorithm was developed for combining tracer data with radar observations to estimate surface circulation. It was assumed that uncertainty of tracer data comes from lack of information on sources and sinks while uncertainty in a radar is due to measuring the radial velocity component only. In that problem we focused on importance of accounting for compatibility of data coming from different sources. Monte Carlo experiments with a 3-vertex system have been carried out to investigate the accuracy of the estimator and to illustrate the role of compatibility.

4. Estimating finite-size Lyapunov exponent by combining real data and model output

In recent years FSLE denoted by λ has become a popular tool for investigating mixing in ocean and atmospheric flows, e.g. [8,9]. In theoretical works the focus was mostly on the scaling of $\lambda(\delta)$ as a function of the initial separation magnitude δ for flows close to isotropic, e.g. [10-12].
To better understand the estimation problem in question theoretical studies are needed to investigate the dependence of FSLE on anisotropy and mixing parameters of mesoscale flows as well as on diffusivity related to submesoscale variability. We focused on a a simplest linear hyperbolic system perturbed by spatially uncorrelated white noise. An explicit solution has been found for a partial differential equation covering the mean separation time of two Lagrangian trajectories. That solution was used to investigate the limit of FSLE as diffusivity indefinitely increases. The case of small diffusivity was addressed as well and hypotheses have been proposed regarding asymptotical behavior of FSLE as diffusivity decays.

5. Further developing theoretical approaches based on fuzzy logic to estimating oceanic parameters from small biased samples.

Sparse observations, biasness, and small samples pose serious obstacles for application of classical statistical methods in processing ocean data. To address these principal challenges we revisited a classical problem in statistics with wide applications in physical oceanography: estimating a location parameter from two different samples. The bottom line in our approach is that the key assumption of unbiased observations is rejected.
An absolute majority of studies in estimating a location parameter addresses, first, linear combinations of either the original sample or its ranking, [13], and, second, unbiased observations. During the reported period we suggested and studied a new class of essentially nonlinear estimators based on the fuzzy set theory ideas [2,3] to handle biased observations coming from two different sources. Because any analytical investigation of the standard error for highly non-linear functions of sample is hard, we concentrated, first, on analytical studying the asymptotical bias of the suggested estimators and, second, on Monte Carlo simulations for small samples with the traditional standard error as an efficiency measure. Different noise distributions were tested including heavy-tailed and that generated by logistic chaos.

RESULTS

1. Twin experiments with a synthetic temperature field showed that the suggested glider/model fusion algorithm is able to reduce both, the bias coming from glider observations due to uncertainties in the glider position and the bias of model caused by a poor resolution. That finding is illustrated

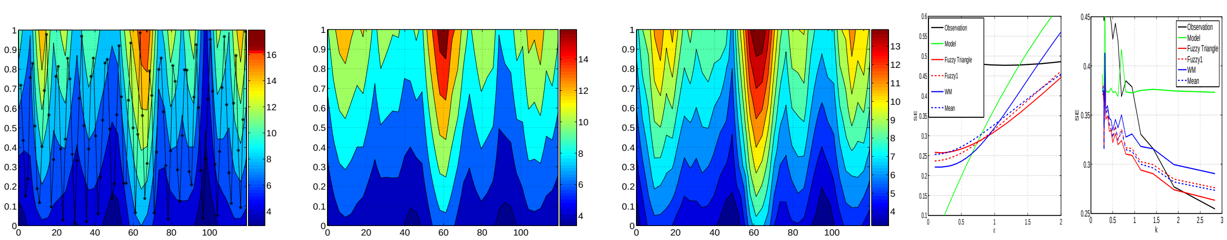

Figure 1. From left to right: 1) 'True' temperature field and glider trajectory. 2) Observed field. 3) Estimated field. 4) Dependence of the estimation error on ε for different estimators : two kinds of WM (blue) and two kinds of FE (red). 5) Dependence of the estimation error on k (right)

in first three panels of Fig. 1 where positions of some anomalies in the estimated field are much

closer to that of the 'true' field than in the observed one (glider 'screening'. At the same time, the estimator captures some submesoscale features which are completely missed in the model output. When investigating the dependence of the estimation error on the submesoscale intensity ε and reduced downcast frequency (or average downcast slope) k for different estimators, we found that, first, the fuzzy estimator (FE) yields a reasonable accuracy in the range $\varepsilon < 2$ and $k > 0.5$, then, FE is better than the weighted mean (WM) for high intensities ε while WM should be preferred for small ε, and finally, both FE and WM decay fast with increasing k and FE is insignificantly better. These conclusions are illustrated in last two panels in Fig. 1

When working on separation of space and time variability for fast changing thermohaline structures we focused on the problem of retrieving the Ligurian front time evolution. Sixteen glider missions across the front have been achieved over a 15-month period between October 2008 and December 2009. For preliminary study, 400 geolocalised profiles acquired during six mission from March 10 through March 22, 2009 were considered (first two panels in Fig.2).

The glider missions were accompanied with in-field efforts in the framework of the Boussole observation program. In particular, the line across the Ligurian front was sampled every month by 7 stations using CTD carousel between 0-400m. Seven profile locations acquired in March 14, 2009 were included in consideration (first panel)

As for methodology we found that a traditional polynomial regression for retrieving the front evolution performed better than two other developed procedure (parametric estimation and fuzzy linear regression). In particular the polynomial regression of glider data showed a perfect agreement with CTD and allowed to estimate the evolution for a longer period of time than other methods.

The results are shown in the last three panels of Fig.2. The recovered time behavior of the front position, curvature, and thickness is in a good agreement with observed wind patterns and with biogeocheimal properties across the front

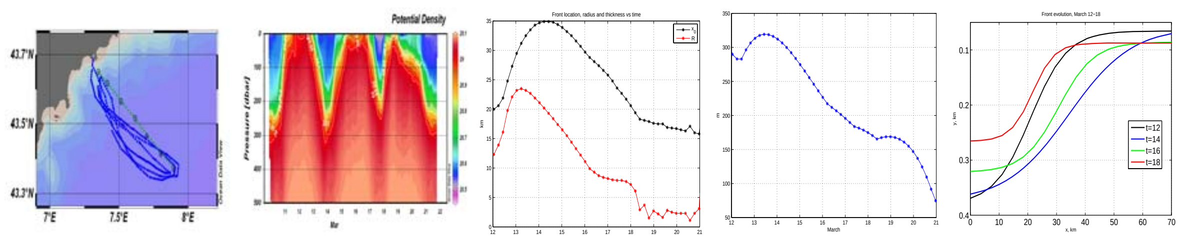

Figure 2. From left to right: 1) Glider profiles and CTD locations. 2) Glider 'screening' of the potential density field. 3) Estimated position (black) and curvature (red) of the front vs time. 4) Estimated front thickness vs time. 5) Evolution of the smoothed front in time.

2. Testing the developed radar/drifter data fusion algorithm in the framework of twin experiments with the regional circulation model (NEMO-GLAzur64, PE z-coordinate, free surface, 1/64deg $\approx$ 1.7km, Gulf of Lions, Ligurian Sea) showed that the estimation error essentially depends on the closeness of the drifters involved to the point where the velocity is estimated. Specifically, reasonable estimates can be obtained at points distanced up to 27 km from available drifter trajectories. Then the procedure was applied to real data (Experiment TOSCA, May 2010, Ligurian Sea, Cape Sicie) with pretty positive results (Fig.3). The obtained circulation is in agreement with the NEMO experiments, however we did not have a chance to compare the results with the real circulation.

3. One of the practical advantages of the developed approach to data fusion is that it allows to introduce a rigorous metrics for quantifying compatibility between two data sets containing information on the same parameter. The main conclusion from experiments with combining radar and tracer data is that accounting for compatibility is of a great importance. We illustrate that

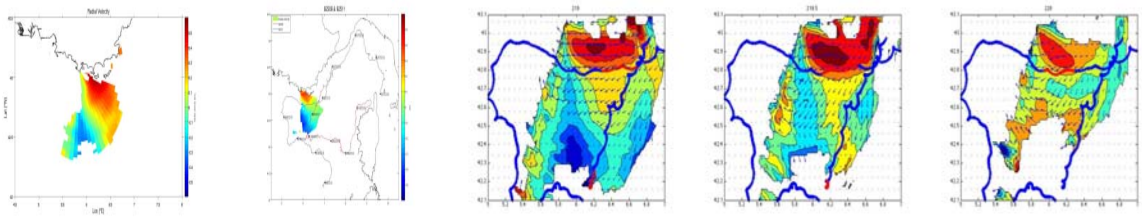

Figure 3. From left to right: Area covered by radar. Drifter trajectories. Estimated circulation for three time moments

finding with the example of 3-vortex system in Fig.4. A version of the algorithm, where the compatibility was quantified and accounted for, turns out to be very efficient in the area of high and medium compatibility (second panel). At the same time, ignoring compatibility leads to a disaster (third panel).

A careful study discovered that at a certain grid point two sources are incompatible if, first, the direction to radar is about orthogonal to tracer lines and, second, the tracer gradients are mostly due to unknown tracer sources rather than to advection. An elaborated error analysis showed that an accurate estimate (error is about 20 %) is provided even if the uncertainty in the forcing and dissipation of tracer is as high as 40% (fourth panel).

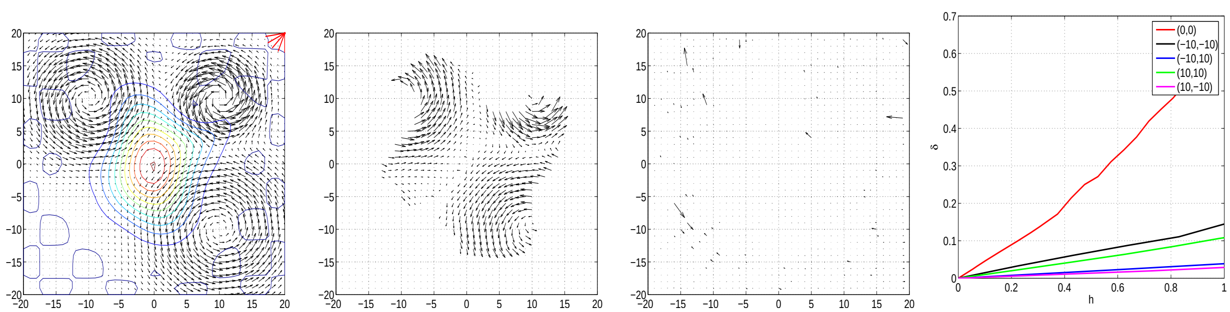

Figure 4. Fusion of radar data and tracer observations. 1) 'True' circulation and observed tracer distribution. 2) Estimated circulation with accounting for compatibility of radar and tracer data 3) Same with no accounting for compatibility. 4) Dependence of the estimation error on uncertainty in tracer forcing and dissipation for four different positions of radar

4. The focus in theoretical studying FSLE denoted by λ was on ability FSLE to detect Lagrangian motion barriers in different dynamical structures superimposed by smaller scale turbulence. During this period we concentrated on a vicinity of a saddle point. Let θ be the direction of initial separation of two Lagrangian particles and D the small scale diffusivity. First, it was shown that there is a little difference between the curves $\lambda = \lambda(\theta)$ for pure dynamics ($D = 0$) and for infinitely large stochastic perturbations ($D = \infty$). A surprising result for small diffusivities was that the limit of λ as D goes to zero differs significantly from λ for the unperturbed dynamics.

In summary, first, our results show that FSLE is an extremely efficient instrument for detecting saddle points of dynamical systems regardless intensity of the stochastic perturbations. Second, the obtained explicit expressions for FSLE in terms of dynamics and turbulence parameters can be used for estimating FSLE from a model output and drifter data.

5. The suggested estimators of the location parameter from two biased samples were compared to the classical least square estimator as well as to other weighted estimators traditionally used in

statistics by two criteria, the asymptotical bias and standard error.

Regarding the asymptotical bias, the considered fuzzy estimators are uniformly better than the classical least square estimator. As for small samples, the new estimators are of higher accuracy than the traditional least square estimator and its modifications for essential bias and high level of noise . Moreover, even for small bias, the fuzzy estimators are only slightly worse than the optimal one.

Unexpectedly, the weighted estimator with equal weights successfully competed with fuzzy estimators for essential biases and modest noise level, but it is of little help for high level of noise or negligible biases.

The results are illustrated in Fig.5 for Cauchy distribution of the noise, but similar conclusions are drawn from other distributions including normal and that generated by logistic chaos.

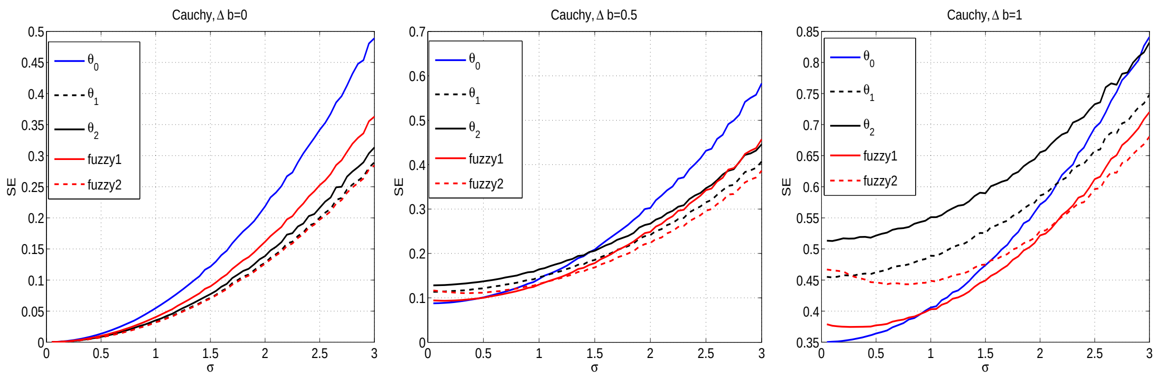

Figure 5. Dependence of the estimation standard error on the noise level σ for different estimators and different values of bias scale Δb (Cauchy noise): weighted median with weights inversely proportional to variances (black solid), weighted median with weights inversely proportional to std (black dashed), weighted median with equal weights (blue), fuzzy estimators based on different membership functions (red). The median is taken as an estimate of center. 1)$\Delta b = 0$; 2) $\Delta b = 0.5$; 3) $\Delta b = 1$;

IMPACT/APPLICATIONS

The developed method for separating space and time variability form glider and CTD observations gives to oceanographers a useful tool for investigating meso- and submesoscale processes in coastal frontal zones.

The proven importance of accounting for compatibility of different data sets on the example of drifter/tracer data fusion could influence the basic principles in combining information coming from different sources. Worthy to stress that compatibility was quantified by a rigorous metrics based on the fuzzy logic approach.

In addition the developed methods are capable to aggregate data at different resolutions and account for sample biasness. Thus, we expect that our results will stimulate more efforts in developing fusion methods which in contrast to traditional assimilation are computationally cheap, portable , and carry no risk of ruining a model during the running time.

Our theoretical findings in studying finite-size Lyapunov exponent provide researchers with efficient tools for detecting saddle points and other barriers for transport. The results also can be used for estimating FSLE by combining a circulation model output and Lagrangian data.

RELATED PROJECTS

"Ocean 3D+", MURI Project, ONR N00014-11-1-0087, PIs: A. Griffa, T. Ozgokmen, I. Mezic, C. Jones , I. Rypina, S. L. Smith, L. Pratt, D. Kirwan

REFERENCES

1. D.Dubious and H. Prade, (1986), Possibility theory, Plenum Press, New York and London,
2. D. Dubois, H.Prade, and R.R. Yager, (1997) Fuzzy Information Engineering, John Wiley & Sons, Inc.
3. G. Shafer,(1976), A mathematical theory of evidence, Princeton University Press.
4. L.I.Piterbarg, (2011), Parameter estimation from small biased samples: statistics vs fuzzy logic, *Fuzzy Sets and Systems*, 170, 1-21
5. Chapman, R.D. and Hans C. Graber, (1997), Validation of HF radar measurements, Oceanography, vol. 10, n.2
6. Yoshikawa, Y., A. Masuda, K. Marubayashi, M. Ishibashi, and A. Okuno (2006), On the accuracy of HF radar measurement in the Tsushima Strait, Journal of Geophysical Reseach, v. 111, C04009, 10 pp. doi:10.1029/2005JC003232
7. Na, H., K.Kim, K.Chang, (2007), Accuracy of surface current velocity measurements obtained from HF radar along the east cost of Korea, http://www.pices.int/publications/presentations/
8. B. Joseph and B. Legras, 2002, Relation between Kinematic Boundaries, Stirring, and Barriers for the Antarctic Polar Vortex, *J. Atm. Sci., v.59, pp.1198-1212*
9. A. Haza, T. Ozgokmen, A. Griffa, Z. Garaffo, L. Piterbarg, (2011), Parameterization of Submesoscale Transport in the Gulf Stream Region Using Lagrangian Subgridscale Models, Ocean Modelling, v.42, 31-49
10. Artale, V., Boffetta, G., Celani, A., Cencini, M., Vulpiani, A., 1997. Dispersion of passive tracers in closed basins: Beyond the diffusion coefficient. Phys. Fluids 9, 31623171.
11. L.I. Piterbarg, (2012), Finite size Lyapunov exponent for some simple models of turbulence, Applied Mathematical Modelling, v.36, n.8, 34643476
12. LaCasce, J. H., (2008), Statistics from Lagrangian observations, *Prog. Oceanogr.*, (77), 1-29.
13. R.J. Huber, Robust estimation of a location parameter, *The Annals of Mathematical Statistics*, 35 (1964) 73-101.

PUBLICATIONS

1. L.I. Piterbarg and L. Ivanov, (2012), Fuzzy-logic based algorithm for estimating circulation patterns, Current Applied Mathematics, v.1, n.1, 17-39
2. L.I. Piterbarg, (2012), Finite size Lyapunov exponent for some simple models of turbulence, Applied Mathematical Modelling, v.36, n.8, 34643476
3. A. Haza, T. Ozgokmen, A. Griffa, Z. Garaffo, L. Piterbarg, (2012), Parameterization of Submesoscale Transport in the Gulf Stream Region Using Lagrangian Subgridscale Models, Ocean Modelling, v.42, 31-49
4. L.I. Piterbarg and L. Ivanov, (2012), Estimating circulation patterns by combining velocity and tracer observations , Open Journal of Applied Science, accepted
5. L.I. Piterbarg, (2012), Estimation of location parameter from two biased samples, Communication in Statistics, submitted
6. L.I. Piterbarg, (2012), Finite size Lyapunov exponent at a saddle point, Physical Review E, submitted